

Where is Anthony Davis Going?

A Case Study by:
Jarrett Brunner and Divya Bhati



The Task:

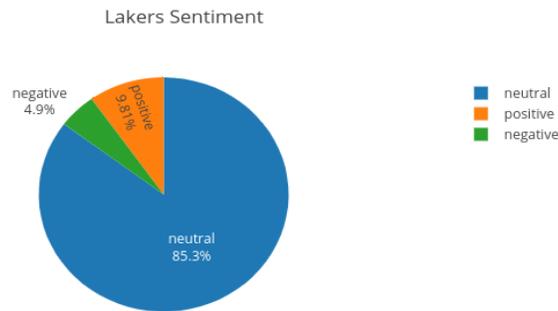
In the days leading up to the NBA trade deadline, the league and its social media outlets had been clouded with discussion regarding where Anthony Davis of the New Orleans Pelicans was headed. Davis has been in the running for Most Valuable Player in the league for multiple years and he's only 25, so he's an extremely valuable asset for virtually every team. He recently announced that he wanted to be traded to a new team, which was followed by an exponential amount of trade rumors and posts on social media. The task here was to try and determine where Twitter users thought Anthony Davis should be traded to based off word sentiment of all the tweets that mention Davis and any of the 30 NBA teams.

Our plan to solve this problem was to do a sentiment analysis using both the [Social Media Macroscope](#) tool and an actual Python script to see the advantages and disadvantages of either approach.

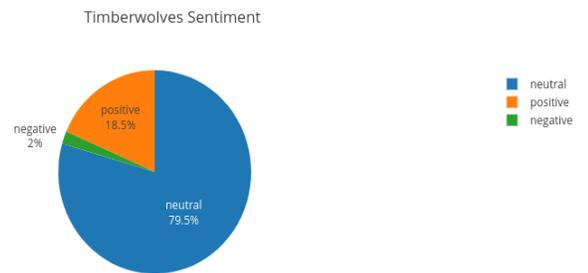
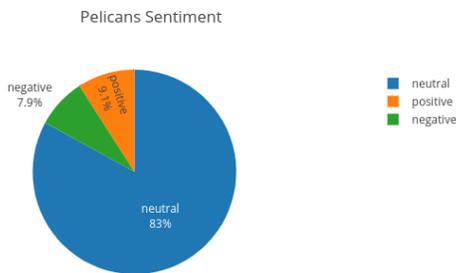
SMM:

We used the Social Media Macroscope's 'SMILE' tool to analyze tweets involving Anthony Davis and each of the 30 NBA teams in order to try to gain some idea of where the public felt he would be traded. To do this, we made thirty separate queries for twitter data by searching "Anthony Davis" and then the name of the NBA team to get a separate dataset of tweets for each team. We then performed Sentiment Analysis on each of these datasets using the VADER tool. After performing the Sentiment Analysis, we saved the circle plots that were generated. We did not do any coding for this part and performed all of this analysis using the Social Media Macroscope itself.

Generally, teams that could viably trade for Anthony Davis had greater percentage of positive sentiment than they did negative sentiment, and teams that could not had a greater negative percentage than positive percentage. However, beyond that, some of the results were surprising. For example, we expected the Los Angeles Lakers to have a strong positive percentage, but they only had about 9% positive sentiment. This is likely because we were unable to categorize and analyze the tweets simultaneously, so we could not weight teams based on the number of tweets about them. In addition, we performed this analysis the day after the Lakers were blown out by the Indiana Pacers, thus contributing more neutral and negative tweets than expected.



Another interesting point is that the New Orleans Pelicans, Anthony Davis' current team, had a strong mix of both positive and negative sentiment. The positive sentiment may be due to people believing that the Pelicans did the right thing in not allowing a team from a larger market to bully them into trading, while the negative sentiment may be from people believing that the Pelicans made the wrong decision to not trade him. One of the most surprising insights we have found is the very strong positive sentiment (18.5%) between Anthony Davis and the Minnesota Timberwolves. Prior to performing this analysis, we had not heard any talk of Davis going to Minnesota. However, after reading some of the tweets from this dataset, we found that there was a lot of talk of Minnesota trading for Davis to partner him with Karl Anthony-Towns.



The Code:

The VADER sentiment tool used by the site is also available as an importable Python library. The first step to analyzing tweet sentiment was to find a data structure that could store tweets sorted by each NBA team. The obvious approach was to use the built-in Python dictionary, which allows you to map a key to a list of values (in this case the key being an NBA team and the value being all the tweets that mention the team). In one loop through our tweet data (stored in AD.csv), we placed each team mention into its respectful slot in the dictionary.

The next step was to perform the actual VADER sentiment analysis. Similar to the first dictionary, we created a second dictionary that would instead store the aggregate sentiment ratings so we can calculate the average rating later. At this point, we also decided to weight the sentiment of each tweet by how many retweets it had, since in the Twitter-verse a retweet is considered its own tweet as well. We then computed the averages by looping through our dictionary and dividing the total sentiment for each team by how many tweets it was mentioned in, and this yielded an average sentiment rating for each team. A basic pseudocode skeleton follows:

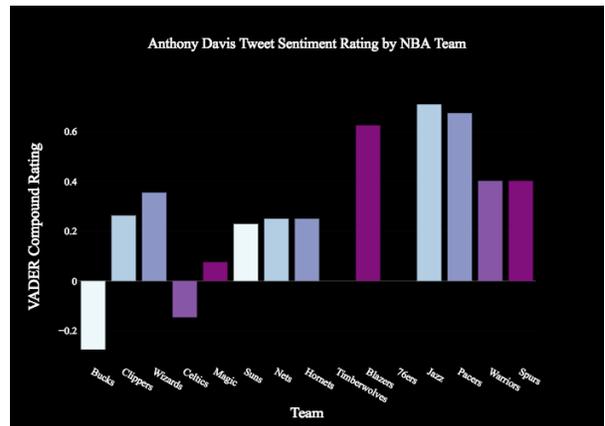
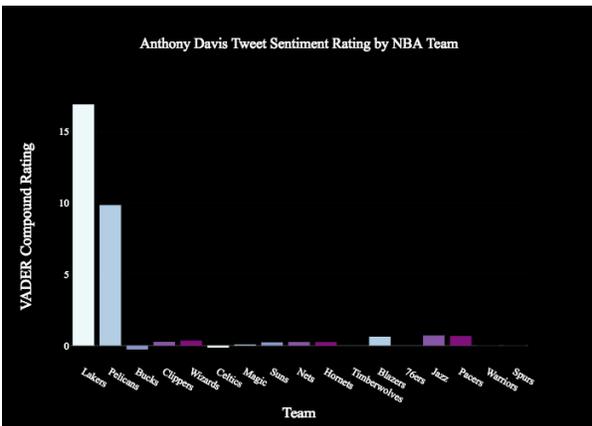
```
#a dictionary that maps each team to all its mentioned tweets
team_tweets = {hawks: "...", celtics: "...", nets: "...", ...}

for each tweet in AD.csv:
    if an NBA team is mentioned:
        add that tweet to our team_tweets dictionary for that specific team

#a dictionary that maps each team to its average sentiment rating
team_sentiments = {hawks: 0, celtics: 0, nets: 0, ...}

for each NBA team:
    total_rating = 0
    for each tweet mentioning that team:
        add its sentiment to its total rating
    average_rating = total_rating / # of tweets
```

Upon executing this code, our “team_sentiments” variable will hold each NBA team mapped to its average sentiment rating per tweet. For simplification, the pseudocode above treats the sentiment rating as a single digit, but really it’s a quadruple of 4 ratings as (positive, neutral, negative, compound). For our purposes, we ended up graphing the compound rating on a bar chart (on the left below).



The Los Angeles Lakers and New Orleans Pelicans are significantly higher than all other teams since the tweets mentioning them were more likely to go viral. We made a second graph (on the right above) that excluded these two teams to better visualize the rest of the NBA. The Milwaukee Bucks and Boston Celtics were the only two teams to have a negative compound rating.